Towards a Formal, Quantitative Molecular Diagnostic Framework

Leslie G. Biesecker
Branch Chief & Senior Investigator
Medical Genomics & Metabolic Genetics Branch
National Human Genome Research Institute
March 14, 2017

The Task

- Determine the pathogenicity of a variant
  - The probability that the variant confers a disease liability
- Make a diagnosis in the patient
  - Use the variant to decide if the patient has the disease

Three Separate Functions

- Critical to distinguish pathogenicity from diagnosis
  - Determine what is known or knowable about the *variant*
    - Clinical laboratory function
  - Use the variant to make a diagnosis (or not)
    - Clinician function
  - Use the diagnosis to change management
    - Clinician function

Nature of the Difficulty

- Highly dimensional problem
- All aspects associated with uncertainty
- Heterogeneity of underlying data
- Utility – Implications
- Values

Nature of the Difficulty

- Highly dimensional problem
  - *Break down into components*
- All aspects associated with uncertainty
  - *Address uncertainty*
- Heterogeneity of underlying data
  - *Weight evidence objectively*
- Utility – Implications
  - *Decouple from utility*
- Values
  - *Preserve professional judgment where appropriate*

The Question That *Will Not Be Discussed*

- What error do you want to make?
- What error will you make without genomics?

Key to Variant Assessment

		Condition (as determined by "Gold standard")		
	Total population	Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$
Test outcome	Test outcome positive	**True positive**	**False positive** (Type I error)	Positive predictive value (PPV, Precision) = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Test outcome positive}}$
	Test outcome negative	**False negative** (Type II error)	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Test outcome negative}}$
	Positive likelihood ratio (**LR+**) = TPR/FPR	True positive rate (TPR, Sensitivity, Recall) = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR, Fall-out) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$

Key to Variant Assessment

		Condition (as determined by "Gold standard")		
	Total population	Condition positive	Condition negative	Prevalence = Σ Condition positive / Σ Total population
Test outcome	Test outcome positive	**True positive**	**False positive** (Type I error)	Positive predictive value (PPV, Precision) = Σ True positive / Σ Test outcome positive
	Test outcome negative	**False negative** (Type II error)	**True negative**	False omission rate (FOR) = Σ False negative / Σ Test outcome negative
	Positive likelihood ratio (**LR+**) = TPR/FPR	True positive rate (TPR, Sensitivity, Recall) = Σ True positive / Σ Condition positive	False positive rate (FPR, Fall-out) = Σ False positive / Σ Condition negative	Accuracy (ACC) = Σ True positive + Σ True negative / Σ Total population

For primary variants higher sensitivity

For secondary variants, higher PPV

Example of Breaking into Components

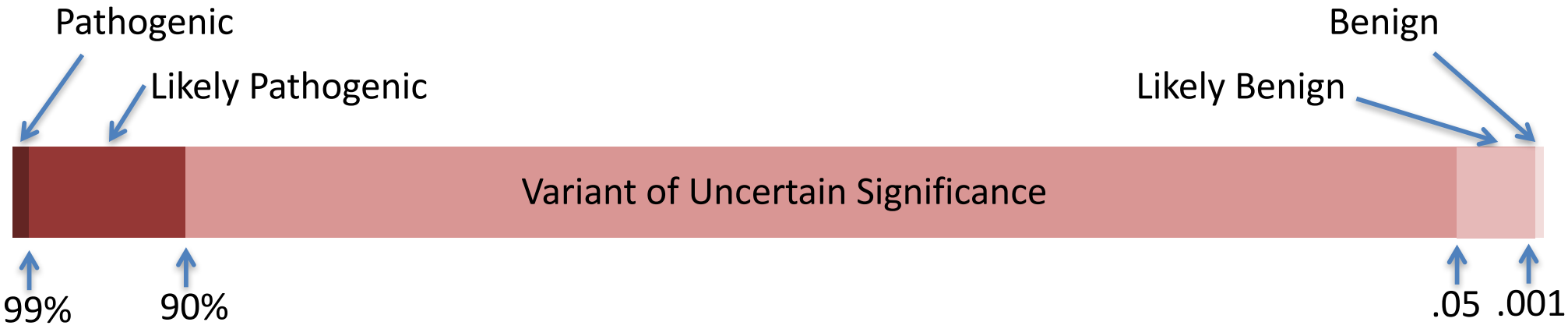
ACMG STANDARDS AND GUIDELINES | Genetics inMedicine

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

General Approach

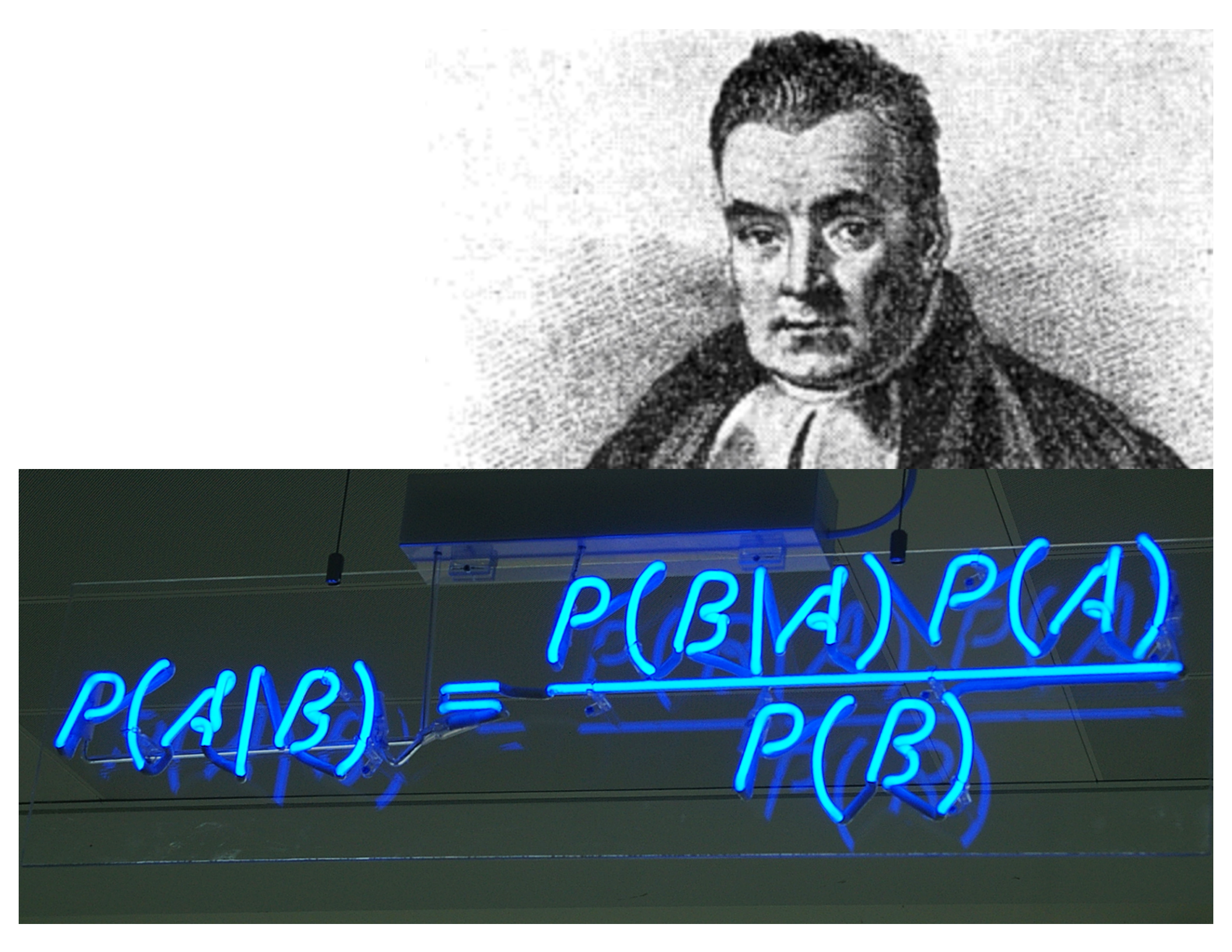
- Adaptation of IARC scale



- Pseudo-quantitative, non-linear, asymmetric assessment of likelihood of pathogenicity

Pathogenicity

- Making real progress
- ACMG Richards et al highly useful
- Can be much better in the future
  - Short, mid, and longer term approaches to make it better


P(A|B) = P(B|A) P(A) / P(B)

Math to English

- $P(A|B) = [P(B|A)*P(A)]/P(B)$

- The probability of A given B equals the probability of B given A times the probability of A all divided by the probability of B

Example: Bean Bags

Bayesian Quantitative Genomics Approach

- Assign variant a prior probability of pathogenicity
  - Dependent on DNA search space
  - *Not* dependent on ascertainment or phenotype
- Then modify this prior based on a piece of evidence
  - Population frequency
  - Bioinformatics
  - Phenotype
  - Etc.

Prior

- Each individual harbors 100 variants that are pathogenic for a Mendelian disorder
- Average person harbors 3×10^6 variants
- Any SNV selected at random, the prior probability that it is pathogenic is $100 / 3 \times 10^6$, or 3.33×10^{-5}

Conditional #1
Variant is in exon or +/- 2 bp

	Pathogenic	Non-Pathogenic
Prior	**3×10^{-5}**	**~1**
Conditional	**0.95***	**0.015****
Joint	**3.16×10^{-5}**	**0.015**
Posterior	**0.0021**	**.9979**

*Estimate that 95% of pathogenic variants for mendelian disorders are in exon or canonical splice bp

**Estimate that 1.5% of genome is exons +/- 2 bp

Conditional #2
Variant is rare

	Pathogenic	Non-Pathogenic
Prior	**0.0021**	**~1**
Conditional	**0.90***	**0.25****
Joint	**0.0019**	**0.25**
Posterior	**0.0075**	**0.9925**

*90% of pathogenic variants are this frequency or rarer
**25% of all variants in genome are this freq or rarer

Etc, etc.

- After all evidence on the variant the posterior probability of pathogenicity is 0.88 (VUS)
- Now what?

Etc, etc.

- After all evidence on the variant the posterior probability of pathogenicity is 0.88 (VUS)
- Now what?
- Look at the patient
- Variant in *PMS2*
- Patient is 44 years old and has had 6 polyps removed + 3 relatives died colon cancer before 60

Conditional #N Phenotype

	Pathogenic	Non-Pathogenic
Prior	**0.88**	**0.12**
Conditional	**0.50***	**0.03****
Joint	**0.44**	**0.0036**
Posterior	**0.992**	**0.008**

*Given pathogenic variant in PMS2 50% patients have this kind of history
*Given no pathogenic variant in PMS2, 3% have this history

A Different Story

- After all evidence on the variant the posterior probability of pathogenicity is 0.88 (VUS)
- Now what?
- Look at the patient
- Variant in *PMS2*
- Patient is 74 years old and has had no polyps or colon cancer

Conditional #N Phenotype

	Pathogenic	Non-Pathogenic
Prior	**0.88**	**0.12**
Conditional	**0.05***	**0.95****
Joint	**0.044**	**0.114**
Posterior	**0.28**	**0.72**

*Given pathogenic variant in PMS2 5% patients have negative history
*Given no pathogenic variant in PMS2, 95% have negative history

Bayesian Quantitative Genomics Approach

- Benefits
  - Separates prior from conditional probabilities
    - Prevents double counting data
    - Facilitates adjusting data
  - Highly amenable to automation
  - Gets us out of “seat of the pants”
  - Uncertainty readily addressed
- Downsides
  - Foreign concept to most clinicians and labs
    - Will require some education
  - We don’t today have most of the needed data

The Future of Genomic Analysis

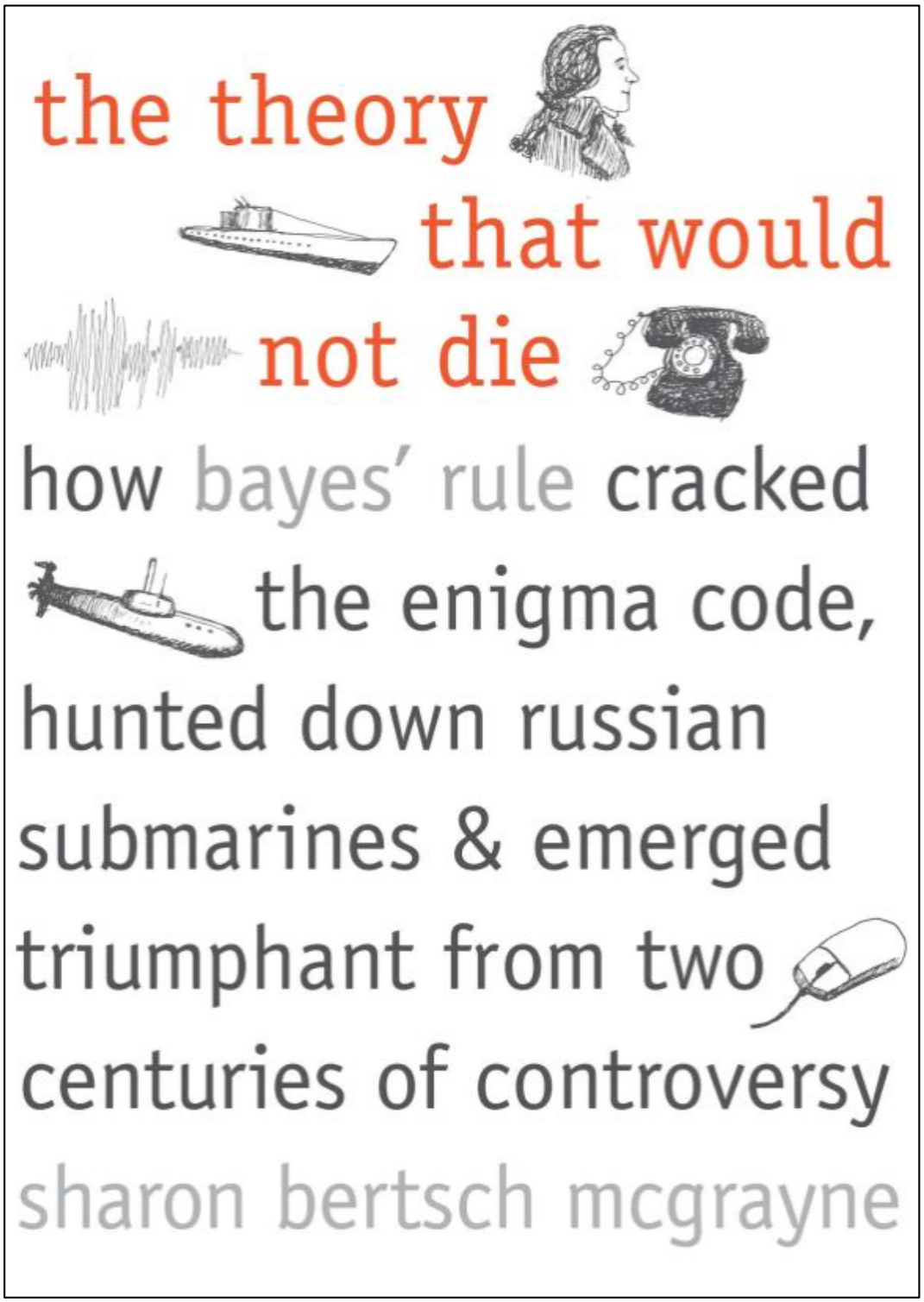
- Separate pathogenicity from diagnosis
- Basic extract of clinical data from EHR to lab
- Sequence
- Semiautomated Bayesian analysis of *every* variant in genome
- CDS tools for interpreting clinicians
  - Post-hoc phenotype driven by genotype supplants pre-hoc phenotype data
- Iterative CDS analyses over lifetime of patient

Will it all be Automated?

“A computer lets you make more mistakes faster than any invention in human history - with the possible exceptions of handguns and tequila.”

— Mitch Ratliff

Read This!



Hat tip: Wendy Rubinstein